(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENTAMT

(12) Übersetzung der **europäischen Patentschrift**

(87) EP 0 557 166 B1

(10) **DE 693 14 514 T 2**

(51) Int. Cl.[6]:
G 10 L 3/02
H 04 R 3/00
G 01 S 3/808

(21) Deutsches Aktenzeichen: 693 14 514.5
(86) Europäisches Aktenzeichen: 93 400 346.8
(86) Europäischer Anmeldetag: 11. 2. 93
(87) Erstveröffentlichung durch das EPA: 25. 8. 93
(87) Veröffentlichungstag der Patenterteilung beim EPA: 15. 10. 97
(47) Veröffentlichungstag im Patentblatt: 12. 2. 98

(30) Unionspriorität:
9201819 18.02.92 FR

(73) Patentinhaber:
Alcatel Alsthom Compagnie Générale d'Electricité, Paris, FR

(74) Vertreter:
Spott Weinmiller & Partner, 82340 Feldafing

(84) Benannte Vertragstaaten:
AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE

(72) Erfinder:
Robbe, Francois, F-95220 Herblay, FR; Dartois, Luc, F-78955 Carrieres Sous Poissy, FR

(54) Rauschverminderungsverfahren in einem Sprachsignal

DE 693 14 514 T 2

93400346.8
Fo 18740 MA/DE

Die Erfindung betrifft ein Verfahren zur Verringerung des akustischen Rauschens in einem Sprachsignal.

Es ist bekannt, daß es günstig ist, bei der Tonaufnahme in verrauschter Umgebung das Umgebungsrauschen zu eliminieren, damit es Aufnahme und Übertragung des Tonsignals unberücksichtigt bleibt. Dies gilt insbesondere im Bereich des Mobilfunks, wo man das Umgebungsrauschen beispielsweise eines Motors des Fahrzeugs, in dem ein Funktelefon verwendet wird, nicht zum Empfänger des Sprachsignals übertragen will.

Der wissenschaftliche Aufsatz von Dal Degan und Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications" (akustische Rauschanalyse und Sprachverstärkungstechniken für Mobilfunkanwendungen), der veröffentlicht wurde in Signal Processing 15, 1988, Seiten 43 bis 56 durch Elsevier Science Publishers B.V., beschreibt und vergleicht verschiedene der Techniken zur Verarbeitung des Rauschens in einem Sprachsignal, das in einem Kraftfahrzeug aufgenommen wurde.

Nach diesem Aufsatz sind Signalverarbeitungsverfahren bekannt, bei denen eine Abschätzung des Spektrums des Umgebungsrauschens erfolgt und dieses Rauschspektrum vom gemessenen Signalspektrum, das von einem Mikrophon kommt, abgezogen wird. Dieses Verfahren, das auf dem Prinzip der Rauschannulierung durch spektrale Subtraktion beruht, wird auch in dem Aufsatz von S.F. Boll "Suppression of acoustic noise in speech using spectral substraction" (Unterdrückung des akustischen Rauschens in einem Sprachsignal unter Verwendung einer spektralen Subtraktion) beschrieben, der veröffentlicht wurde in IEEE Trans. ASSP. Vol.ASSP-27, 1979, Seiten 113 bis 120.

Der Hauptnachteil dieser Methode besteht jedoch darin, daß das Rauschspektrum oft aktualisiert werden muß,

um die Veränderungen des Umgebungsrauschens zu berücksichtigen, und daß diese Aktualisierung nur erfolgen kann, solange der Benutzer nicht spricht, d.h. während der Schweigeperioden. In einer Umgebung, in der das Umgebungsrauschen oft und erheblich variiert, insbesondere in einem Kraftfahrzeug, braucht man also zahlreiche Schweigeperioden, um das Spektrum des Umgebungsrauschens häufig zu aktualisieren. Man verfügt jedoch nicht stets über ausreichend lange Schweigeperioden, um das Spektrum des Rauschens zu aktualisieren, so daß, wenn der Abstand zwischen Schweigeperioden zu groß wird, das Rauschspektrum sich verschlechtert und kurze Rauscherscheinungen nicht mehr berücksichtigen kann. Die Qualität der übertragenen Sprachsignale wird dadurch beeinträchtigt.

Ziel der Erfindung ist es insbesondere, diese Nachteile zu beheben.

Außerdem ist aus der Druckschrift US-A-4 653 102 ein Verfahren zur Verringerung des Umgebungsrauschens in einem Sprachsignal bekannt, das ein Doppelmikrophon, eine Berechnung der schnellen Fouriertransformierten bezüglich der empfangenen Signale und eine Frequenzverarbeitung verwendet. Die akustische Aufnahme einer akustischen Rauschquelle in einem Doppelmikrophon aufgrund einer Berechnung der Phasendifferenz zwischen den von den beiden Mikrophonen aufgefangenen Signalen ist auch bekannt aus der Druckschrift US-A-4 333 170.

Genauer betrachtet ist Ziel der Erfindung ein Verfahren zur Verarbeitung von akustischen Signalen, durch das das Umgebungsrauschen erheblich gedämpft und somit die Qualität einer Sprachsignalübertragung erhöht werden kann, wobei diese Dämpfung des Rauschens ausgehend von einem Rauschspektrum erfolgt, das aktualisiert wird, ohne daß Schweigeperioden auf Seiten des Sprechers erforderlich wären.

Dieses Ziel sowie weitere Ziele, die nachfolgend

aufscheinen, werden erreicht durch ein Verfahren zur Verringerung des akustischen Rauschens in einem empfangenen akustischen Signal, das ein Sprachsignal enthält, wobei die spektralen Rauschkomponenten von dem empfangenen akustischen Signal abgezogen werden, um das Spektrum des Sprachsignals zu bilden, dadurch gekennzeichnet, daß es darin besteht,

- zyklisch durch die Fouriertransformierte digitale Signale, die aus der Analog-Digital-Umwandlung von durch zwei Mikrophone, die einen festen Abstand voneinander haben und das akustische Signal empfangen, gelieferten Signalen in zwei Serien von diskreten Angaben zu verwandeln, wobei jede diskrete Angabe der Serien für die Energie und die Phase eines Frequenzbands des Spektrums des akustischen Signals repräsentativ ist und die Frequenzbänder aneinander anschließen sowie für das Spektrum des empfangenen akustischen Signals repräsentativ sind;
- den dominanten Ankunftswinkel des empfangenen akustischen Signals ausgehend von den Phasenunterschieden zwischen den diskreten Angaben entsprechend den gleichen Frequenzbändern dieser Serien zu bestimmen, wobei der dominante Ankunftswinkel dem Ankunftswinkel des Sprachsignals entspricht,
- ein augenblickliches Spektrum des empfangenen akustischen Signals entsprechend einer der Serien von diskreten Angaben oder entsprechend einer Kombination der diskreten Serien zu bilden, woraus sich eine Verstärkung des Sprachsignals bezüglich des Rauschens ergibt,
- ein Rauschspektrum zu aktualisieren, indem für jedes Frequenzband des augenblicklichen Spektrums der Absolutwert der Differenz zwischen dem dominanten Ankunftswinkel und dem Ankunftswinkel des betreffenden Frequenzbands mit einem Toleranzschwellwert verglichen wird, wobei das Rauschspektrum von den gleichen Frequenzbändern wie das augenblickliche Spektrum gebildet wird und die Energien der Frequenzbänder des Rauschspektrums mit Hilfe der Energien der Frequenzbänder des augenblicklichen Spektrums aktualisiert

werden, deren Absolutwert der Differenz zwischen dem dominanten Ankunftswinkel und dem Ankunftswinkel dieser Frequenzbänder größer als der Toleranzschwellwert ist;
- das aktualisierte Rauschspektrum von dem augenblicklichen Spektrum abzuziehen, um ein Ausgangsspektrum zu erhalten, das von dem Spektrum des Sprachsignals gebildet wird.

Das Prinzip dieses Verfahrens beruht daher auf der Bewertung eines dominanten Ankunftswinkels entsprechend der Lage des Sprechers bezüglich der beiden Mikrophone, die das Tonsignal empfangen, um das Sprachsignal und das Rauschsignal durch spektrale Subtraktion zu trennen.

Vorzugsweise enthält das Verfahren auch einen Schritt der Korrektur des aktualisierten Rauschspektrums abhängig vom Ergebnis der Subtraktion.

Vorzugsweise besteht die Korrektur des aktualisierten Rauschspektrums darin, nach der Subtraktion die Anzahl von Frequenzbändern zu zählen, deren Energie einen Energieschwellwert übersteigt, und das aktualisierte Rauschspektrum durch die Gesamtheit des augenblicklichen Spektrums zu ersetzen, wenn die Anzahl der Frequenzbänder, deren Energie den Energieschwellwert übersteigt, geringer als ein vorgegebener Zahlenwert ist.

Vorzugsweise werden die den Energieschwellwert übersteigenden Energien der diskreten Angaben des Ergebnisses der Subtraktion auf Null gesetzt, ehe das Ergebnis das Ausgangsspektrum bildet.

Auf diese Weise werden die Rauschträgerfrequenzen großer Amplitude im Ausgangsspektrum eliminiert.

Vorzugsweise und gemäß einer ergänzenden Ausführungsform besteht die Korrektur des aktualisierten Rauschspektrums darin, nach der Subtraktion die Anzahl von Frequenzbändern zu zählen, deren Energie den Energieschwellwert übersteigt, und die Bänder des aktualisierten Rauschspektrums durch die Bänder des augenblicklichen Spektrums zu ersetzen, die ein negatives Ergebnis nach der Subtraktion

erbracht haben, wenn die Anzahl der Frequenzbänder, in denen die Energie den Energieschwellwert übersteigt, größer als der vorgegebene Zahlenwert ist.

Gemäß einer bevorzugten Ausführungsform erfolgt die Bestimmung des dominanten Ankunftswinkels dadurch, daß in Speicherfächern für jedes der Frequenzbänder die Gewichte proportional zur Energie der Frequenzbänder summiert werden, wobei jedes Speicherfach einem Intervall des Ankunftswinkels entspricht und die Gewichte in den Speicherfächern entsprechend den Ankunftswinkel der Frequenzbänder summiert werden, und daß der dominante Ankunftswinkel dem Ankunftswinkel entspricht, der dem Speicherfach mit dem größten Gewicht zugewiesen ist.

So wird eine Bestimmung des dominanten Ankunftswinkels entsprechend der Lage des Sprechers bezüglich der beiden Mikrophone durchgeführt, wenn dieser Sprecher ein Sprachsignal liefert.

Vorzugsweise sind die Gewichte auch proportional zu den Frequenzen der Frequenzbänder und die Summierungen bestehen darin, gleitende Mittelwerte zu bilden.

Gemäß einer bevorzugten Ausführungsform besteht die Kombination der diskreten Serien, mit der eine Verstärkung des Sprachsignals bezüglich des Rauschsignals erhalten wird, darin,

- die diskreten Angaben einer der Serien mit denen der anderen Serie ausgehend vom dominanten Ankunftswinkel in Phase zu bringen, so daß die diskreten Angaben der Serien in Phase kommen, deren Ankunftswinkel dem dominanten Ankunftswinkel entspricht;
- die diskreten Angaben der in Phase gebrachten Serien zu summieren, um die diskreten Angaben entsprechend dem Sprachsignal gegenüber den diskreten Angaben entsprechend dem akustischen Rauschen zu verstärken.

Das erfindungsgemäße Verfahren wird vorzugsweise bei der Verarbeitung eines Sprachsignals in einem Mobiltelefon

angewendet.

Andere Merkmale und Vorzüge der Erfindung werden nun anhand eines Ausführungsbeispiels des erfindungsgemäßen Verfahrens beschrieben, das die Erfindung nicht einschränken soll und in einer Vorrichtung durchgeführt wird, deren Blockdiagramm in der einzigen beiliegenden Figur dargestellt ist.

Das Verfahren gemäß der vorliegenden Erfindung kann in sieben aufeinanderfolgende Schritte zerlegt werden, die nachfolgend erläutert werden:

- ein erster Schritt der Signalverarbeitung besteht darin, die von den beiden ortsfesten Mikrophonen gelieferten Signale zu digitalisieren und Fouriertransformierte dieser digitalen Signale zu bilden, um zwei Serien von diskreten Angaben zu erhalten, wobei jede diskrete Angabe einer Serie repräsentativ für die Energie und die Phase eines bestimmten Frequenzbands des von den beiden Mikrophonen empfangene akustischen Signals ist;

- ein zweiter Verarbeitungsschritt besteht darin, den Ankunftswinkel des von den beiden ortsfesten Mikrophonen empfangenen Signals ausgehend von den zwischen zwei identischen Frequenzbändern der diskreten Serien existierenden Phasendifferenzen zu bestimmen. Die Kenntnis dieses Ankunftswinkels gibt Auskunft über die Lage des Sprechers bezüglich der beiden Mikrophone;

- ein dritter Verarbeitungsschritt besteht darin, die von den beiden ortsfesten Mikrophonen gelieferten Sprachsignale phasenrichtig zu rekombinieren, um die Leistung des Sprachsignals im Vergleich zu der des Rauschens zu erhöhen;

- ein vierter Verarbeitungsschritt besteht darin, das Rauschspektrum ausgehend vom Ankunftswinkel des Sprachsignals zu aktualisieren;

- ein fünfter Verarbeitungsschritt besteht darin, das Rauschspektrum vom augenblicklichen gemessenen Spektrum abzuziehen, um ein Ausgangsspektrum zu erhalten;

- ein sechster Verarbeitungsschritt besteht darin, das Rauschspektrum abhängig vom Ergebnis der Subtraktion zu korrigieren;

- ein siebter und letzter Verarbeitungsschritt besteht darin, das Ausgangssignal zu rekonstruieren, um beispielsweise seine Aussendung zu erlauben (Anwendung auf das Mobiltelefon).

Der erste Schritt der Signalverarbeitung besteht darin, die von zwei Mikrophonen gelieferten analogen Signale zu digitalisieren und auf diese digitalen Signale eine Fouriertransformation anzuwenden, um zwei Serien von digitalen Signalen zu erhalten.

Die Mikrophone 10 und 11 (siehe beiliegende Figur) bilden eine akustische Antenne. Sie sind beide ortsfest, und die Vorrichtung, die die Erfindung verwendet, ist vorzugsweise auf ein Freisprechsystem angewendet. Die von den Mikrophonen 10 und 11 kommenden analogen Signale werden an Analog-Digital-Wandler 12 bzw. 13 angelegt, die auch eine Filterung der Signale (Frequenzband 300 bis 3400 Hz) durchführen. Die digitalisierten Signale gelangen dann in Hamming-Fenster 14, 15 in digitaler Form und werden anschließend in digitale Vektoren durch Vorrichtungen 16 und 17 zur Berechnung von schnellen Fouriertransformierten (FFT) des Rangs N umgewandelt.

Die Analog-Digital-Wandler 12 und 13 liefern nacheinander je 256 digitale Werte an einen nicht dargestellten Ausgangsspeicher mit 512 Speicherplätzen. In einem Zeitpunkt t liefert der Speicher 512 digitale Werte an ein Hamming-Fenster, die von den 256 im Zeitpunkt t berechneten Werten hinter 256 weiteren Werten gebildet werden, die zum Zeitpunkt t-1 berechnet wurden. Jedes Hamming-Fenster 14 und 15 führt zu einer Dämpfung der Sekundärkeulen des empfangenen Signals und ermöglicht die Erhöhung der Signalauflösung durch die FFT 16 und 17. Jedes dieser Fenster 14 und 15 liefert 512 digitale Werte an eine der Vorrichtungen 16 und

17. Letztere liefern je eine Serie von 512 Frequenzbändern, die die Frequenzen von 0 bis 8 kHz untereinander aufteilen. Jedes Frequenzband mißt also etwas mehr als 15 Hz. Ein Taktgeber H wird in den FFT-Rechenvorrichtungen 16 und 17 verwendet.

Dieser Taktgeber kann auch durch eine Vorrichtung zum Zählen der Anzahl von durch die FFT 16 und 17 gelieferten Tastproben ersetzt werden, die auf Null gesetzt wird, wenn 512 Verarbeitungen durchgeführt wurden.

Die Ergebnisse S1 und S2 dieser Berechnungen werden dann von einer Folge von Vektoren in digitaler Form gebildet, wobei jeder Vektor einer Tastprobe des Spektrums eines der Eingangssignale entspricht und beispielsweise aus zwei Wörtern zu je 16 Bits besteht, die eine komplexe Zahl definieren.

In Wirklichkeit können nur 256 unterschiedliche Vektoren berücksichtigt werden, da im Fall eines (im mathematischen Sinn) reellen Signals der Absolutwert der Fouriertransformierten eine geradzahlige Funktion ist, während die Phase ungerade ist. Die 256 verbleibenden Vektoren werden nicht berücksichtigt. Am Ausgang des FFT liefert jeder Kanal also 256 Vektoren, die je aus zwei Wörtern von 16 Bits bestehen.

Der zweite Verarbeitungsschritt besteht darin, den Ankunftswinkel des Signals zu bestimmen, das von den beiden ortsfesten Mikrophonen aufgefangen wurde, und zwar aufgrund der Phasendifferenzen zwischen zwei identischen Frequenzkanälen der diskreten Serien.

Die aus den FFT kommenden Vektoren werden nacheinander an eine Vorrichtung 18 zur Berechnung der Phasenverschiebung zwischen den von den Mikrophonen 10 und 11 kommenden Signalen geliefert. Die Serien S1 und S2 sind durch identische Frequenzbänder charakterisiert, wobei jedem Frequenzband jedes Eingangskanals eine Phase $\Phi 1$ und $\Phi 2$ sowie unterschiedliche Absolutwerte entsprechen, wenn das vom

Mikrophon 10 ankommende akustische Signal nicht genau dem gleicht, das vom Mikrophon 11 kommt (Phasenverschiebung der Signale aufgrund der Laufzeitunterschiede).

Die Signale S1 und S2 bestehen also aus Serien von Vektoren, wobei jedes Vektorpaar einem Frequenzband entspricht. Diese Signale werden an eine Vorrichtung 18 zur Berechnung der zwischen den Signalen S1 und S2 existierenden Phasenverschiebung geliefert, und zwar Frequenzband für Frequenzband.

Da der Abstand zwischen den beiden Mikrophonen bekannt ist und das akustische Signal einer ebenen Welle angenähert werden kann, gewinnt man den Ankunftswinkel des akustischen Signals in jedem Frequenzband aus der folgenden Gleichung:

$$\sin \Theta = v \cdot \delta\Phi / (2 \cdot \pi \cdot d \cdot f)$$

Hierbei ist:

- Θ der gesuchte Ankunftswinkel des akustischen Signals im betrachteten Frequenzband,
- v die Schallgeschwindigkeit,
- $\delta\Phi$ die Phasendifferenz zwischen den beiden Signalen,
- d der Abstand zwischen den Mikrophonen 10 und 11,
- f die Frequenz in Hertz entsprechend dem betreffenden Frequenzband. Die Frequenz f ist beispielsweise die zentrale Frequenz dieses Bands.

Die Rechenvorrichtung 18 ermittelt also die Phasenverschiebung zwischen den aus den beiden Mikrophonen kommenden Signalen für jedes Frequenzband.

Die Vorrichtung 18 liefert die Ankunftswinkel Θ, die für die verschiedenen Frequenzbänder berechnet wurden, an eine Vorrichtung 19 zur Suche des Ankunftswinkels. Die Vorrichtung 19 aktualisiert ein Histogramm der Ankunftswinkel mit m Fächern, die den Winkelbereich von -90 bis +90° umfassen. Jedes Fach hat daher eine Länge von 180/m Grad.

Die Aktualisierung des Histogramms besteht für jedes Frequenzband darin, in dem dem durch die Rechenvorrichtung

18 berechneten Ankunftswinkel entsprechenden Fach ein Gewicht proportional zur Frequenz und proportional zur Energie hinzuzufügen, die in dem betreffenden Frequenzband vorhanden ist (Amplitude der spektralen Komponente). Dieses Zugewicht ist vorzugsweise proportional zur Frequenz, da die Bestimmung des Winkels bei höheren Frequenzen zuverlässiger ist, die sowohl eine bessere Annäherung an eine ebene Welle als auch eine geringere Frequenzabweichung $\delta f/f$ aufweisen. Der in einem Fach gespeicherte Wert ist dann das Ergebnis eines gleitenden Mittelwerts, der aus der folgenden Gleichung berechnet wird:

$$c(n) = a \cdot c(n-1) + (1-a) \cdot poids(n)$$

Hierbei ist:

- c(n) der in einem Fach des Histogramms in dem Zeitpunkt n vorliegende Wert;
- a eine reelle Zahl kleiner als 1 und nahe bei 1;
- poids(n) der Wert des Gewichts zum Zeitpunkt n. Dieser Wert ist beispielsweise gleich der Energie des betreffenden Bands multipliziert mit der Frequenz dieses Bands.

Wenn in dem von den beiden Mikrophonen 10 und 11 empfangenen akustischen Signal kein Sprachsignal enthalten ist, nehmen die in den verschiedenen Speicherfächern gespeicherten Werte mit jeder neuen Angabe ab, so daß schließlich die Gewichte der verschiedenen Speicherfächer einander im wesentlichen gleichen, denn das Rauschen verteilt sich gleichförmig auf die verschiedenen Fächer, wenn es nicht von einer lokalisierten Quelle wie z.B. dem Motor eines Fahrzeugs stammt.

Das Histogramm wird also periodisch aktualisiert, beispielsweise alle 32 ms.

Nicht dargestellte Mittel bieten außerdem die Möglichkeit, die Aktualisierung des Histogramms zu blockieren, wenn ein Sprachsignal vom Benutzer empfangen wird.

Wenn die Aktualisierung in allen Frequenzbändern

durchgeführt wurde, sucht die Vorrichtung 17 das Maximum des Histogramms, d.h. das Fach mit dem größten Gewicht. Die Lage dieses Fachs, d.h. der Ankunftswinkel, der ihm zugewiesen ist, entspricht dem dominanten Ankunftswinkel. Dieser dominante Ankunftswinkel Θ_{max} ist derjenige, unter dem die akustischen Signale mit der größten Energie ankommen.

Wenn das akustische Signal einen Sprachanteil enthält, entspricht der dominante Ankunftswinkel Θ_{max} der Lage des Sprechers bezüglich der beiden Mikrophone. Die Rauschfrequenzen, die von nicht lokalisierten Quellen stammen, verteilen sich nämlich nahezu gleichmäßig auf die verschiedenen Fächer des Histogramms, während die Sprachsignale von einer lokalisierten Quelle (einem Sprecher) sich stets in dem gleichen Fach akkumulieren und rasch eine Spitze in dem Fach des Histogramms erscheinen lassen, die dem dominanten Ankunftswinkel Θ_{max} entspricht.

Gemäß einer anderen Ausführungsform erzeugt man ebensoviele Diagramme wie es Frequenzkanäle gibt und bildet einen Mittelwert der verschiedenen Diagramme über alle Kanäle, um den dominanten Ankunftswinkel zu ermitteln. Diese Ausführungsform erfordert jedoch größere Speicherkapazitäten, so daß es besser ist, gleitende Mittelwerte für jedes Speicherfach zu berechnen.

Andere Ausführungsformen, mit denen der Ankunftswinkel des Sprachsignals erkannt werden kann, können auch eingesetzt werden.

Der zweite Signalverarbeitungsschritt des erfindungsgemäßen Verfahrens führt also zur Kenntnis des Ankunftswinkels Θ_{max} der Sprachsignale.

Der dritte Signalverarbeitungsschritt besteht darin, die von den beiden Mikrophonen gelieferten Signale phasenrichtig zu kombinieren. Diese Kombination dient einer Verstärkung des Sprachsignals bezüglich des Rauschsignals. Dieser Schritt erfolgt in den Blöcken 20 und 21, bei denen es sich um Mittel zur Phasenanpassung der beiden Kanäle und

zur phasenrichtigen Addition der Kanalsignale handelt.

Die Vorrichtung 19 zur Suche des Ankunftswinkels liefert den Mitteln 20 zur Phasenanpassung den Wert Θ_{max} des dominanten Ankunftswinkels. Die Mittel 20 berechnen für jedes Frequenzband die Phasendifferenz zwischen den beiden Eingangskanälen für den dominanten Ankunftswinkel Θ_{max}, der ihnen von der Vorrichtung 19 geliefert wird. Diese Berechnung erfolgt ausgehend von der obigen Gleichung, in der Θ durch Θ_{max} ersetzt wird, also:

$$\delta\Phi = (2 \cdot \pi \cdot d \cdot f \cdot \sin\Theta_{max})/v$$

Die für jedes Frequenzband erhaltene Phasendifferenz wird mit der Phase eines der beiden Signale summiert (oder von ihr abgezogen, je nach der Art der Berechnung von $\delta\Phi$). In der dargestellten Ausführungsform wird die Phasendifferenz zu S2 hinzuaddiert (oder davon abgezogen). Die Mittel 20 zur Phasenverschiebung ergeben also ein Signal S2, dessen Frequenzbänder entsprechend dem Sprachsignal mit denen des Signals S1 in Phase sind, da diese Frequenzbänder Träger des größten Energieanteils sind, durch die Θ_{max} bestimmt werden kann.

Dann bilden die Mittel 21 zur phasenrichtigen Addition der Kanäle die Summe des Signals S1 und des phasenrichtigen Signals S2. Indem die Signale der beiden Kanäle phasenrichtig addiert werden, summiert sich das Sprachsignal kohärent und man erhält ein Sprachsignal einer größeren Amplitude. Dagegen ist das Rauschsignal aufgrund von dessen spektraler Verteilung abgeschwächt (das Rauschsignal stammt nicht von einer lokalisierten Quelle wie das Sprachsignal). Die phasenrichtige Summierung der Signale ergibt somit eine Verstärkung des Sprachsignals gegenüber dem Rauschsignal.

Die Eliminierung des Rauschens ist jedoch im allgemeinen nicht ausreichend, und es verbleibt ein Restrauschen, dessen spektrale Komponenten denselben Ankunftswinkel wie das Sprachsignal haben. Eine zusätzliche Verarbeitung ist daher notwendig.

Es sei bemerkt, daß dieser dritte Schritt der Signalverarbeitung fakultativ ist und daß eines der beiden Signale, beispielsweise das Signal S2, unmittelbar für den weiteren Fortgang des Verfahrens benutzt werden kann. In diesem Fall wird das aus den Summiermitteln 21 kommende Signal durch das Signal S2 ersetzt.

Es ist auch möglich, eine größere Anzahl von fest installierten Mikrophonen zu verwenden. Die Verwendung von digitalen Signalen entsprechend Frequenzanteilen von aus mehreren Mikrophonen stammenden Signalen macht den Rechenalgorithmus für den Ankunftswinkel Θ für jedes Frequenzband jedoch kompliziert, ebenso wie die Phasendrehung dieser Signale, um beispielsweise ihre Summierung zu erlauben. Es ist auch nicht interessant, die Signale entsprechend dem von einem dritten Mikrophon kommenden akustischen Signal zu verwenden, damit diese Signale die aus dem Summierer 21 ersetzen, da ihre Ankunftswinkel zwingend sich von denen der Signale S1 und S2 unterscheiden und der gefundene dominante Ankunftswinkel die von diesem dritten Mikrophon stammenden Signale nicht berücksichtigen könnte.

Der zusätzliche Verfahrensschritt bildet den vierten oben angegebenen Schritt und besteht darin, das Rauschspektrum zu aktualisieren. Diese Aktualisierung des Rauschspektrums erfolgt einerseits ausgehend vom Ankunftswinkel Θ_{max}, der als derjenige des Sprachsignals erkannt wurde, und andererseits ausgehend vom augenblicklichen Spektrum, das aus den Serien von digitalen Angaben gebildet wird, die von den Additionsmitteln geliefert wurden.

Für jedes Frequenzband vergleicht eine Vorrichtung 22 zur Aktualisierung des Rauschspektrums den von der Rechenvorrichtung 18 berechneten Ankunftswinkel Θ mit dem Ankunftswinkel Θ_{max} des Sprachsignals, der von der Vorrichtung 19 geliefert wurde. Die Vorrichtung 22 kann beispielsweise den Absolutwert der Differenz zwischen Θ_{max} und Θ für jedes Frequenzband mit einer Toleranzschwelle

Θ_s vergleichen.

Wenn der Absolutwert dieser Abweichung zwischen den beiden Winkeln größer als die Toleranzschwelle Θ_s ist, wird das entsprechende Frequenzband als zum Rauschspektrum gehörend betrachtet. Die in diesem Frequenzband vorhandene Energie wird dann zu Aktualisierung des Rauschspektrums, beispielsweise durch gleitende Mittelwertbildung, verwendet. Diese Aktualisierung kann natürlich auch darin bestehen, einfach einen Teil der Angaben des Rauschspektrums durch die entsprechenden Angaben des augenblicklichen Rauschspektrums zu ersetzen. Dieses Rauschspektrum wird in einem digitalen Speicher 23 gespeichert. Durch die Einführung einer Toleranzschwelle Θ_s kann man kleine Veränderungen der Lage des Sprechers bezüglich der beiden Mikrophone und auch Rechenungenauigkeiten berücksichtigen.

Wenn der Absolutwert der Abweichung zwischen den beiden Winkeln Θ_{max} und Θ kleiner als die Toleranzschwelle Θ_s ist, wird das betrachtete Frequenzband als zum Sprachspektrum gehörend betrachtet, und seine Energie wird nicht zur Aktualisierung des Rauschspektrums im Speicher 23 herangezogen.

Die Vorrichtung 22 ermöglicht also eine Aktualisierung des Rauschspektrums durch Vergleich des Ankunftswinkels Θ jedes Frequenzbands mit dem dominanten Ankunftswinkel Θ_{max}. Dieser berechnete Winkel Θ_{max} hat also die Aufgabe, eine Auswahl der Frequenzen des von den FFT erhaltenen Spektrums zu erlauben. Natürlich ist es nicht unbedingt erforderlich, einen Absolutwert der Differenz zwischen dem dominanten Ankunftswinkel und dem Ankunftswinkel jedes Frequenzbands des augenblicklichen Spektrums zu ermitteln. Es ist beispielsweise auch möglich, einen Ankunftswinkelbereich einer Breite $2\Theta_s$ zu begrenzen, der auf Θ_{max} zentriert ist, und zu überprüfen, ob der Ankunftswinkel Θ jedes Frequenzbands sich in diesem Bereich befindet.

Es sei bemerkt, daß die Aktualisierung des Rausch-

spektrums kontinuierlich alle 32 ms erfolgt, ob nun ein Sprachanteil im von den beiden Mikrophonen 10 und 11 empfangenen Signal vorhanden ist oder nicht. Das erfindungsgemäße Verfahren unterscheidet sich also vom obengenannten Stand der Technik dadurch, daß keine Schweigeperioden erforderlich sind, um das Rauschspektrum aktualisieren zu können, denn die Bestimmung, ob ein Frequenzband zum Rauschspektrum oder zum Sprachspektrum gehört, erfolgt ausgehend vom berechneten dominanten Ankunftswinkel und dem Ankunftswinkel für das betreffende Frequenzband.

Der fünfte Schritt besteht darin, das Rauschspektrum vom gemessenen augenblicklichen Spektrum abzuziehen.

Dieser Schritt erfolgt in einer Vorrichtung 24 zur Subtraktion des Rauschspektrums vom augenblicklichen Spektrum. Das Rauschspektrum wird aus dem digitalen Speicher 23 ausgelesen und vom augenblicklichen Spektrum abgezogen, das von der Vorrichtung 22 kommt. Wenn die Verstärkung gemäß dem zweiten Schritt des vorliegenden Verfahrens entfällt wird das augenblickliche Spektrum durch die Vektoren eines der beiden Signale, beispielsweise des Signals S2, gebildet.

Diese Subtraktion ergibt ein Ausgangsspektrum, das von einem praktisch vollständig von spektralen Rauschkomponenten befreiten Sprachsignalspektrum gebildet wird. Es ist jedoch möglich, eine weitere Verarbeitung des erhaltenen Spektrums durchzuführen, mit dem eine Korrektur des aktualisierten Rauschspektrums möglich ist.

Nach der Subtraktion werden die negativen Ergebnisse auf Null gesetzt. Dann können sich zwei Fälle ergeben:

- Das augenblickliche Spektrum enthält kein Sprachsignal, und das Restspektrum enthält dann nur eine kleine Anzahl von signifikanten Frequenzen.

- Das augenblickliche Spektrum enthält ein Sprachsignal, so daß das verbleibende Spektrum eine große Zahl von energiehaltigen Frequenzen entsprechend im wesentlichen dem Sprachspektrum enthält.

Um den Inhalt des augenblicklichen Spektrums zu kennen, braucht man also nur die Anzahl von Frequenzbändern zu zählen, für die die spektrale Leistung über einem Schwellwert S_p liegt, der es erlaubt, die Frequenzbänder mit geringem Energieanteil unberücksichtigt zu lassen und zu eliminieren. Diese Frequenzbänder entsprechen entweder dem Restrauschen oder Sprachfrequenzbändern, aber ihre Energie ist so gering, daß sie nicht zum Empfänger übertragen werden müssen (Anwendung für das Mobiltelefon).

Vorzugsweise ist der Schwellwert Sp nicht für jeden Frequenzband der gleiche und hängt von der Energie ab, die in jedem dieser Frequenzbänder vorliegt. Man kann beispielsweise einen ersten Schwellwert den Frequenzbänder zwischen 0 und 2 kHz und einen zweiten Schwellwert, beispielsweise halb so groß wie der erste Schwellwert, den Frequenzbändern zwischen 2 und 4 Khz zuweisen. Dadurch kann man die Tatsache berücksichtigen, daß die Energie des Rauschspektrums bei niedrigen Frequenzen in einem Fahrzeug größer als bei höheren Frequenzen ist.

Wenn die Anzahl von Frequenzbändern mit einer Energie größer als Sp gering ist (geringer als ein zahlenmäßiger Schwellwert), werden die gezählten Frequenzbänder als Rauschrestfrequenz enthaltend betrachtet. Die Gesamtheit des augenblicklichen Spektrums (d.h. die am Eingang der Vorrichtung 22 vorliegenden Angaben) wird dann zur Aktualisierung des Rauschspektrums verwendet. Diese Operation findet im digitalen Speicher 23 statt und bildet den sechsten Verarbeitungsschritt des Signals. Er besteht genauer betrachtet darin, die Energie der Frequenzbänder des aktualisierten Rauschspektrums durch die Energie der entsprechenden Frequenzbänder des augenblicklichen Spektrums zu ersetzen. Außerdem werden die Frequenzbänder, deren Energie den Schwellwert Sp übersteigt, vor dem Ersatz der Frequenzbänder des Rauschspektrums auf Null gesetzt. So werden energiereiche Rauschfrequenzen großer Amplitude eliminiert.

In einer Ausführungsvariante werden nur die Frequenzbänder des augenblicklichen Spektrums mit größeren Energieanteilen als die der entsprechenden Frequenzbänder des Rauschspektrums für den Ersatz herangezogen. So berücksichtigt man ausschließlich Frequenzbänder des augenblicklichen Spektrums mit einem großen Energieanteil.

Wenn das nach der Subtraktion erhaltene Spektrum einem Sprachsignal entspricht (d.h. daß die Anzahl von Frequenzbändern mit einem Energieanteil größer als Sp nach der Subtraktion größer als der digitale Schwellwert ist, werden nur die Energieanteile der Frequenzbänder des augenblicklichen Spektrums entsprechend den Frequenzbändern des restlichen Spektrums nach der Subtraktion, die ein negatives Ergebnis enthalten, zur Korrektur des Rauschspektrums verwendet. Ein negatives Ergebnis nach der Subtraktion bedeutet nämlich, daß das entsprechende Frequenzband des aktualisierten Rauschspektrums eine zu große Energie besitzt. Diese Korrektur hilft zu vermeiden, daß das restliche Rauschspektrum, d.h. das aktualisierte Spektrum, nur noch aus einigen Frequenzbändern großer Amplitude besteht, was dieses Spektrum besonders unangenehm und für die Rekonstruktion der Sprache störend machen würde.

Natürlich ist die Korrektur des Rauschspektrums gemäß diesem sechsten Verfahrensschritt fakultativ und kann auf unterschiedliche Arten realisiert werden, sofern man sich entscheidet, ob das durch die Subtraktion erhaltene Spektrum als ein Spektrum betrachtet werden soll oder nicht, das weiterzuverarbeitende Sprachfrequenzbänder enthält, beispielsweise zur Übertragung an eine Zielperson.

Der siebte und letzte Verarbeitungsschritt besteht darin, ein analoges Ausgangssignal zu bilden, das beispielsweise ausgesendet werden soll. Dieser Schritt verwendet eine Vorrichtung 25 zur Erzeugung des Ausgangssignals mit einer Vorrichtung 26 für die inverse schnelle Fouriertransformierte (FFT^{-1}), die 512 Sprachtastproben liefert. Der FFT^{-1}-Vor-

richtung geht eine nicht dargestellte Vorrichtung voraus, die die 256 empfangenen Vektoren regenerieren kann, um 512 Eingangsvektoren für die FFT^{-1}-Vorrichtung zu erhalten. Auf die Vorrichtung 26 erfolgt eine Überdeckungsvorrichtung 27, die eine einfache Rekonstruktion des Ausgangssignals erlaubt. Die Vorrichtung 27 überdeckt die 256 ersten empfangenen Tastproben mit den 256 letzten Tastproben, die vorher empfangen wurden (also vorher verarbeitet wurden). Diese Überdeckung erlaubt eine ausgangsseitige Kompensation der Anwendung eines Hamming-Fensters am Eingang. Ein Digital-Analog-Wandler 28 ergibt ein wenig verrauschtes akustisches Signal, das für eine Aussendung zu einer Zielperson bereit ist. Man kann dieses Signal auch beispielsweise auf einem Magnetband speichern oder einer anderen Verarbeitung zuführen.

Dieser siebte Schritt kann in manchen Anwendungen entfallen. Beispielsweise kann das erfindungsgemäße Verfahren zur Spracherkennung verwendet werden, wobei in diesem Fall der siebte Verarbeitungsschritt entfällt, da die Spracherkennungsvorrichtungen die spektrale Darstellung eines Sprachsignals weiter auswerten.

Das erfindungsgemäße Verfahren erlaubt also eine deutliche Verringerung des Rauschspektrums in einem akustischen Signal und liefert ein Sprachsignal, ohne Schweigeperioden des Sprechers für die Aktualisierung des Rauschspektrums zu benötigen, da die Erkennung des Ankunftswinkels des Signals für die Trennung zwischen Rauschen und Sprache herangezogen wird.

93400346.8

ANSPRÜCHE

1. Verfahren zur Verringerung des akustischen Rauschens in einem empfangenen akustischen Signal, das ein Sprachsignal enthält, wobei die spektralen Rauschkomponenten von dem empfangenen akustischen Signal abgezogen werden, um das Spektrum des Sprachsignals zu bilden, dadurch gekennzeichnet, daß es darin besteht,

- zyklisch durch die Fouriertransformierte (16, 17) digitale Signale, die aus der Analog-Digital-Umwandlung (12, 13) von durch zwei Mikrophone (10, 11), die einen festen Abstand voneinander haben und das akustische Signal empfangen, gelieferten Signalen in zwei Serien (S1, S2) von diskreten Angaben zu verwandeln, wobei jede diskrete Angabe der Serien für die Energie und die Phase eines Frequenzbands des Spektrums des akustischen Signals repräsentativ ist und die Frequenzbänder aneinander anschließen sowie für das Spektrum des empfangenen akustischen Signals repräsentativ sind;

- den dominanten Ankunftswinkel (Θ_{max}) des empfangenen akustischen Signals unter den berechneten Ankunftswinkeln Θ ausgehend von den Phasenunterschieden zwischen den diskreten Angaben entsprechend den gleichen Frequenzbändern dieser Serien (S1, S2) zu bestimmen (18, 19), wobei der dominante Ankunftswinkel (Θ_{max}) dem Ankunftswinkel des Sprachsignals bezüglich der beiden Mikrophone entspricht,

- ein augenblickliches Spektrum des empfangenen akustischen Signals entsprechend einer der Serien (S1, S2) von diskreten Angaben oder entsprechend einer Kombination (20, 21) der diskreten Serien (S1, S2) zu bilden, woraus sich eine Verstärkung des Sprachsignals bezüglich des Rauschens ergibt,

- ein Rauschspektrum (23) zu aktualisieren (22), indem für jedes Frequenzband des augenblicklichen Spektrums der Absolutwert der Differenz zwischen dem dominanten Ankunfts-

winkel (Θ_{max}) und dem Ankunftswinkel (Θ) des betreffenden Frequenzbands mit einem Toleranzschwellwert (Θ_s) verglichen wird, wobei das Rauschspektrum (23) von den gleichen Frequenzbändern wie das augenblickliche Spektrum gebildet wird und die Energien der Frequenzbänder des Rauschspektrums (32) mit Hilfe der Energien der Frequenzbänder des augenblicklichen Spektrums aktualisiert werden, deren Absolutwert der Differenz zwischen dem dominanten Ankunftswinkel (Θ_{max}) und dem Ankunftswinkel (Θ) dieser Frequenzbänder größer als der Toleranzschwellwert (Θ_s) ist;

- das aktualisierte Rauschspektrum (23) von dem augenblicklichen Spektrum abzuziehen (24), um ein Ausgangsspektrum zu erhalten, das von dem Spektrum des Sprachsignals gebildet wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß es darin besteht, das aktualisierte Rauschspektrum (23) abhängig vom Ergebnis der Subtraktion (24) zu korrigieren.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, daß die Korrektur des aktualisierten Rauschspektrums (23) darin besteht, nach der Subtraktion die Anzahl von Frequenzbändern zu zählen, deren Energie einen Energieschwellwert (Sp) übersteigt, und das aktualisierte Rauschspektrum (23) durch die Gesamtheit des augenblicklichen Spektrums zu ersetzen, wenn die Anzahl der Frequenzbänder, deren Energie den Energieschwellwert (Sp) übersteigt, geringer als ein vorgegebener Zahlenwert ist.

4. Verfahren nach Anspruch 3, dadurch gekennzeichnet, daß die den Energieschwellwert (Sp) übersteigende Energie der diskreten Angaben des Ergebnisses der Subtraktion auf Null gesetzt wird, ehe das Ergebnis das Ausgangsspektrum ersetzt.

5. Verfahren nach einem der Ansprüche 2 bis 4, dadurch

gekennzeichnet, daß die Korrektur des aktualisierten Rauschspektrums (23) darin besteht, nach der Subtraktion die Anzahl von Frequenzbändern zu zählen, deren Energie den Energieschwellwert (Sp) übersteigt, und die Bänder des aktualisierten Rauschspektrums (23) durch die Bänder des augenblicklichen Spektrums zu ersetzen, die ein negatives Ergebnis nach der Subtraktion (242) erbracht haben, wenn die Anzahl der Frequenzbänder, in denen die Energie den Energieschwellwert (Sp) übersteigt, größer als der vorgegebene Zahlenwert ist.

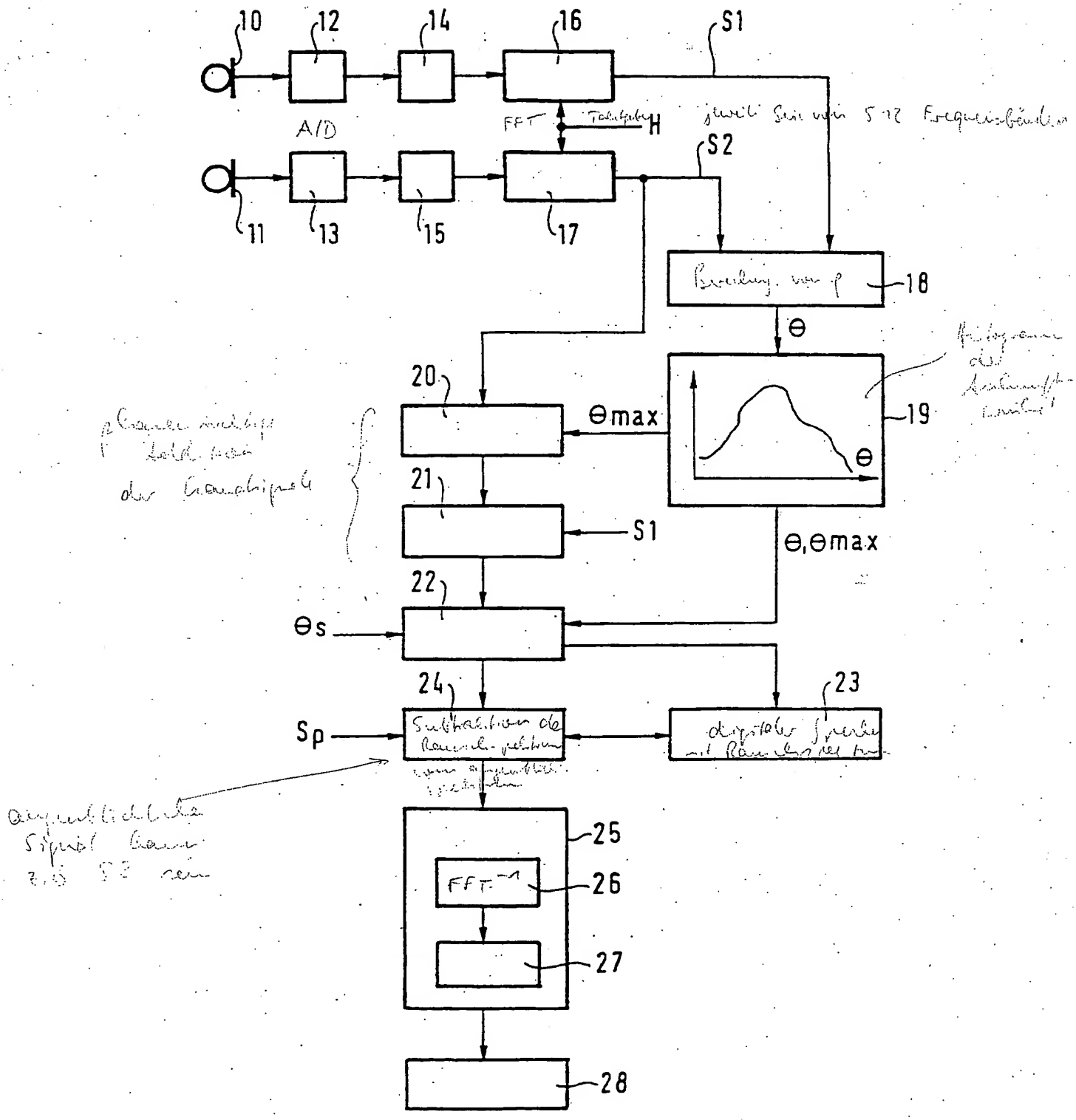
6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß die Bestimmung (18, 19) des dominanten Ankunftswinkels (Θ_{max}) dadurch erreicht wird, daß in Speicherfächern für jedes der Frequenzbänder die Gewichte proportional zur Energie der Frequenzbänder summiert werden, wobei jedes Speicherfach einem Intervall des Ankunftswinkels entspricht und die Gewichte in den Speicherfächern entsprechend den Ankunftswinkel (Θ) der Frequenzbänder summiert werden, und daß der dominante Ankunftswinkel (Θ_{max}) dem Ankunftswinkel (Θ) entspricht, der dem Speicherfach mit dem größten Gewicht zugewiesen ist.

7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, daß die Gewichte auch proportional zu den Frequenzen der Frequenzbänder sind.

8. Verfahren nach einem der Ansprüche 6 und 7, dadurch gekennzeichnet, daß die Summierungen darin bestehen, gleitende Mittelwerte zu bilden.

9. Verfahren nach einem der Ansprüche 1 bis 8, dadurch gekennzeichnet, daß die Kombination (20, 21) der diskreten Serien (S1, S2), mit der eine Verstärkung des sprachsignals bezüglich des Rauschsignals erhalten wird, darin besteht,

93400346.8
0 557 166



- die diskreten Angaben einer der Serien (S2) mit denen der anderen Serie (S1) ausgehend vom dominanten Ankunftswinkel (Θ_{max}) in Phase zu bringen (20), so daß die diskreten Angaben der Serien (S1, S2) in Phase kommen, deren Ankunftswinkel dem dominanten Ankunftswinkel (Θ_{max}) entspricht;
- die diskreten Angaben der in Phase gebrachten Serien (S1, S2) zu summieren (21), um die diskreten Angaben entsprechend dem Sprachsignal gegenüber den diskreten Angaben entsprechend dem akustischen Rauschen zu verstärken.

10. Verfahren nach einem der Ansprüche 1 bis 9, dadurch gekennzeichnet, daß es auf die Verarbeitung eines Sprachsignals in einem Mobiltelefon angewendet wird.